

# Fast simulations for intensity mapping experiments

David Alonso,<sup>1★</sup> Pedro G. Ferreira<sup>1</sup> and Mario G. Santos<sup>2,3,4</sup>

<sup>1</sup>*Sub-department of Astrophysics, Department of Physics, University of Oxford, DWB, Keble Road, Oxford OX1 3RH, UK*

<sup>2</sup>*Department of Physics, University of Western Cape, Cape Town 7535, South Africa*

<sup>3</sup>*SKA SA, 3rd Floor, The Park, Park Road, Pinelands 7405, South Africa*

<sup>4</sup>*CENTRA, Instituto Superior Técnico, Universidade de Lisboa, Lisboa P-1049-001, Portugal*

Accepted 2014 August 12. Received 2014 August 8; in original form 2014 May 27

## ABSTRACT

We present a code to generate mock observations of 21 cm intensity mapping experiments. The emphasis of the code is on reducing the computational cost of running a full-blown simulation, trading computational time for accuracy. The code can be used to generate independent realizations of the cosmological signal and foregrounds, which are necessary, for instance, in order to obtain realistic forecasts for future intensity mapping experiments. The code is able to reproduce the correct angular and radial clustering pattern for the cosmological signal, including redshift-space distortions, lightcone evolution and bias. Furthermore, it is possible to simulate a variety of foregrounds, including the potentially problematic polarized synchrotron emission.

**Key words:** large-scale structure of Universe – radio lines: galaxies.

## 1 INTRODUCTION

The distribution of perturbations in the matter density supplies a wealth of information on the late-time evolution of the Universe on cosmological scales. Until now, such studies have focused on optical redshift surveys (Colless et al. 2001; Jones et al. 2004; Drinkwater et al. 2010; Anderson et al. 2014), which are able to trace the galaxy distribution to redshifts  $z \lesssim 1.5$  on relatively large scales, providing constraints on multiple observational probes, such as baryon acoustic oscillations (BAO) or redshift-space distortions (RSDs). One of the challenges of this type of observations is the large integration times that are necessary in order to obtain a good determination of the galaxy redshifts from their optical spectra. This limits the number density of sources with observed redshifts and restricts the maximum radial distance that can be reached. A possible alternative solution to this problem is to use only the photometry of each object to obtain a fast but inaccurate redshift measurement. Photometric redshift surveys (Kaiser et al. 2002; Tyson 2002; Abbott et al. 2005) can thus observe a much higher number density of sources and reach larger redshifts (limited only by the sensitivity of each particular experiment), at the cost of losing almost all the relevant information in the radial direction ( $\delta r_{\parallel} \sim \mathcal{O}(100 \text{ Mpc } h^{-1})$ ).

A promising alternative has recently been proposed in the context of radio astronomy. This approach, labelled *intensity mapping* (Battye, Davies & Weller 2004; Barkana & Loeb 2005; McQuinn et al. 2006; Chang et al. 2008; Loeb & Wyithe 2008; Mao et al. 2008; Pritchard & Loeb 2008; Wyithe & Loeb 2008; Wyithe, Loeb

& Geil 2008; Peterson et al. 2009; Bagla, Khandai & Datta 2010; Seo et al. 2010; Lidz et al. 2011; Ansari et al. 2012; Battye et al. 2013; Bull et al. 2014), is based on measuring the radio emission from different patches of the sky and different frequencies. Any pocket of neutral hydrogen (H I) will emit in the isolated 21 cm line, corresponding to the hyperfine spin-flip transition of the 1s ground state. This is observed at a redshifted frequency  $\nu_{\text{obs}} = \nu_{21}/(1+z)$ , where  $\nu_{21} = 1420.4 \text{ MHz}$  is the rest-frame frequency of the 21 cm line. Thus, by measuring the intensity of radio emission from different directions in the sky, it is in principle possible to trace the three-dimensional distribution of H I in the Universe. The aim of intensity mapping is therefore not to focus on measuring the flux of individual galaxies, but rather the combined emission arriving from relatively wide patches of the sky. Even though this technique misses any information related to the very small scale perturbations, it is possible to make very fast 3D maps of the H I density on large scales, which are the most relevant ones for cosmological studies.

Analysing the data from any large-scale structure experiment requires extensive use of simulated catalogues to estimate statistical uncertainties, study potential sources of systematic errors, assess the model independence of the results, etc. On the one hand, a large number of independent realizations are usually needed [e.g.  $\mathcal{O}(100\text{--}1000)$  for a good convergence of the covariance matrix in galaxy surveys (Manera et al. 2013)]. On the other hand, it is desirable to have mock catalogues which describe the observed field as accurately as possible. The computational cost of generating these mocks usually grows with the level of precision required by the experiment, and therefore a compromise must be reached between computational speed and accuracy. In particular, since 21 cm intensity mapping can access large redshifts ( $z \gtrsim 2\text{--}3$ ),  $N$ -body

★E-mail: david.alonso@astro.ox.ac.uk

simulations covering these volumes ( $\sim 500 \text{ Gpc}^3$ ) with a reasonable mass resolution ( $\sim 10^{11} M_\odot$ ) can be an expensive solution. Along these lines, a lot of work has been done within the community of galaxy redshift surveys to develop alternative cheaper methods that can still reproduce the relevant physical observables (de la Torre & Peacock 2013; Tashev, Zaldarriaga & Eisenstein 2013; White, Tinker & McBride 2014).

On the other hand, one of the most important challenges facing H I intensity mapping is the presence of foregrounds (both galactic and extragalactic) with amplitudes several orders of magnitude larger than the signal to be measured. The statistical properties, as well as the frequency dependence of these foregrounds, differ significantly from those of the signal, and therefore there is hope that they can be successfully subtracted (Di Matteo et al. 2002; Oh & Mack 2003; Santos, Cooray & Knox 2005; Morales, Bowman & Hewitt 2006; Wang et al. 2006; Gleser, Nusser & Benson 2008; Jelić et al. 2008; Bernardi et al. 2009, 2010; Liu et al. 2009; Jelić et al. 2010; Moore et al. 2013; Shaw et al. 2013; Wolz et al. 2014). Nevertheless, this foreground subtraction is a potential source of systematic effects that could limit the observational power of intensity mapping for cosmology. Evaluating and modelling these systematics is therefore an essential step in the observational pipeline that requires the use of simulated realizations of these foregrounds.

It has become the norm in the analysis of cosmic microwave background (CMB) data to construct efficient simulations which can then be used to understand the analysis pipeline for any given experiment (Hinshaw et al. 2013; Planck Collaboration 2013a). By including different foreground contaminants and instrumental systematic effects in the simulation, it is then possible, via Monte Carlo techniques, to accurately estimate the various biases that may enter the final result. It is in this spirit that we approach the problem of generating mock realizations of the 21 cm emission by H I after reionization ( $z \lesssim 4$ ), as well as its most relevant foregrounds. The method we present is similar to those used by several galaxy redshift surveys, and is based on generating a lognormal realization of the density field of H I. Through this method, it is possible to implement different important effects (e.g. the bias of H I with respect to the matter density, the lightcone evolution of the density field, RSDs, frequency decorrelation in the foregrounds, etc.) at a very low computational cost.

Section 2 reviews the key theoretical results on which the method used to generate the cosmological signal is based, as well as the theoretical models used to validate the simulations. The actual method is explained in Section 3. The models and procedures used to generate the foreground simulations are described in Section 4. We study the validity of these simulations in Section 5. Finally, the main results of this study are summarized in Section 6.

## 2 THEORETICAL BACKGROUND

### 2.1 Linear perturbation theory

The growth of structure via gravitational collapse is a non-linear process which, even in an expanding background, quickly departs significantly from the solutions of the linearized equations (Peebles 1980). However, linear perturbation theory is an extremely useful tool when studying structure formation, since it contains vital information which can be used to understand the late-time distribution of the matter density field. Furthermore, since, as we will see in the next section, the lognormal transformation is able to describe the

distribution of the late-time non-linear density field in terms of the linear one, linear perturbation theory is, in our case, essential.

In the Newtonian limit, the linearized equation that describes the evolution of the overdensity field  $\delta$  is given by

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2} \frac{\Omega_M}{a^2} \delta = 0,$$

where the derivatives  $()$  are taken with respect to the cosmic time  $t$ ,  $\Omega_M$  is the non-relativistic matter parameter and  $H \equiv \dot{a}/a$  is the expansion rate. Note that this equation is valid both in Fourier and real space, since the evolution of linear perturbations in the dust component is self-similar. The growing-mode solution to this equation is

$$\delta \propto g(a) \equiv \frac{5}{2} \Omega_M H_0^2 H(a) \int_0^a \frac{da'}{[a' H(a')]^3}, \quad (1)$$

where  $H_0$  is the local Hubble rate and we have used the normalization  $g(a \ll 1) \simeq a$ . We will denote the growth factor normalized to its present value by  $D(a) \equiv g(a)/g(1)$ . Thus, at the linear level  $\delta(\mathbf{x}, a) = \delta_0(\mathbf{x}) D(a)$ , where  $\mathbf{x}$  are comoving coordinates. Since we want to generate mock catalogues in redshift space, we are also interested in the peculiar velocity field, which causes RSDs. For most interesting cases, the peculiar velocity field can be modelled as being irrotational to a good approximation (vector modes are predicted to decay rapidly), and therefore can be written as the gradient of a velocity potential  $\mathbf{v} \simeq \nabla \varphi$ . The velocity field is coupled to the overdensity field through the continuity equation, which reads

$$a f(a) H(a) \delta_k - k^2 \varphi_k = 0 \quad (2)$$

for a dust-like component, where  $f(a) \equiv d \ln D / d \ln a$  is the growth rate. The redshift distortion is then given by the radial comoving peculiar velocity

$$\Delta z_{\text{RSD}} = \frac{v_r}{c a} = \hat{\mathbf{u}}_r \cdot \frac{\nabla \varphi}{a}. \quad (3)$$

For our purposes, the most useful result from linear perturbation theory for a dust component is that both the overdensity and the peculiar velocity fields can be self-similarly related to their present-day values

$$\delta(\mathbf{x}, z) = D(z) \delta_0(\mathbf{x}), \quad \mathbf{v}(\mathbf{x}, z) = \frac{f(z) H(z) D(z)}{(1+z) H_0 f_0} \mathbf{v}_0(\mathbf{x}). \quad (4)$$

These first-order perturbations are linearly related to the primordial perturbations, which seem to be Gaussianly distributed (Planck Collaboration 2013b). Therefore, it is possible to produce realizations of the Fourier-space overdensity field by generating Gaussianly distributed random numbers with a variance given by the power spectrum

$$\langle \delta_{k_1} \delta_{k_2} \rangle \equiv \delta(\mathbf{k}_1 + \mathbf{k}_2) P(\mathbf{k}). \quad (5)$$

However, a Gaussian overdensity field cannot represent a realistic matter density, since, for one thing, values of  $\delta < -1$  should have a probability of exactly 0.

### 2.2 Lognormal fields

Lognormal fields were first proposed and analysed by Coles & Jones (1991) as a possible model to describe the distribution of matter in the Universe. A lognormal random field is defined in terms of a Gaussian random field through the local transformation

$$x_{\text{LN}} = \exp x_{\text{G}}. \quad (6)$$

One of the nice properties of these fields is that, while the Gaussian variable  $x_G$  is allowed to take any values in  $(-\infty, +\infty)$ ,  $x_{LN}$  can only take positive values by construction. Furthermore, as discussed in Coles & Jones (1991), the density field evolved along Lagrangian trajectories according to the linear velocity field can be well described by a lognormal distribution, which justifies the use of lognormal fields from a physical point of view. In order to obtain a lognormal overdensity field with zero mean from a Gaussian field, the transformation (6) must be slightly varied, the correct equation being

$$1 + \delta_{LN} = \exp\left(\delta_G - \frac{\sigma_G^2}{2}\right), \quad (7)$$

where  $\sigma_G^2$  is the variance of the Gaussian overdensity field.

Lognormal density fields have been used in the past by different collaborations to create fast mock galaxy catalogues using techniques similar to the ones described in Section 3 (Cole et al. 2005; Beutler et al. 2011; Blake et al. 2011), and are, therefore, a well-established tool. The simulation of the extragalactic radio sky by Wilman et al. (2008) was also based on a similar method.

Nevertheless, the lognormal transformation cannot be expected to describe the galaxy density correctly on arbitrarily small scales ( $\lesssim 5 \text{ Mpc } h^{-1}$ ; Kitaura, Jasche & Metcalf 2010), and therefore this technique is only useful for a limited number of analyses.

### 2.3 The intensity mapping power spectrum

The intensity in an observed frequency bin  $\delta\nu$  coming from the 21 cm emission of an object at redshift  $z$  with H I mass  $M_{H\text{I}}$ , subtending a solid angle  $\delta\Omega$ , is given by (Abdalla & Rawlings 2005)

$$I(\nu, \hat{n}) = \frac{3 h_p A_{12}}{16\pi m_H} \frac{1}{((1+z)r(z))^2} \frac{M_{H\text{I}}}{\delta\nu \delta\Omega} \nu_{21}, \quad (8)$$

where  $A_{12} = 2.876 \times 10^{-15} \text{ Hz}$  is the Einstein coefficient corresponding to the emission from the 21 cm hyperfine transition,  $h_p$  is the Planck constant and  $m_H = 1.6733 \times 10^{-24} \text{ g}$  is the hydrogen atom mass. Here,  $r(z)$  is the comoving curvature distance  $r(z) = c \text{ sinn}(H_0 \sqrt{|\Omega_k|} \chi(z)/c)/(H_0 \sqrt{|\Omega_k|})^1$  and  $\chi(z)$  is the radial comoving distance

$$\chi(z) = \int_0^z \frac{c dz'}{H(z')}. \quad (9)$$

This intensity  $I(\nu, \hat{n})$  can be written in terms of a blackbody temperature in the Rayleigh–Jeans approximation  $T = I c^2/(2 k_B \nu^2)$ , where  $k_B$  is Boltzmann’s constant. Using this we can estimate the mean brightness temperature coming from redshift  $z$  and its fluctuations in terms of the H I density:

$$T_{21}(z, \hat{n}) = (0.19055 \text{ K}) \frac{\Omega_b h (1+z)^2 x_{H\text{I}}(z)}{\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}} (1 + \delta_{H\text{I}}). \quad (10)$$

Here  $x_{H\text{I}}(z)$  is the H I fraction (i.e. fraction of the total baryon density in H I) and  $\delta_{H\text{I}}$  is the H I overdensity field in redshift space (smoothed over the volume defined by  $\delta\nu$  and  $\delta\Omega$ ).

<sup>1</sup> We use the usual convention where  $\text{sinn}(x)$  is  $\sin(x)$ ,  $\sinh(x)$  and  $x$  for closed, open and flat universes, respectively.

#### 2.3.1 Angular clustering

For a given frequency band, the brightness temperature can be transformed into harmonic space

$$a_{lm}(\nu) = \int d\hat{n}^2 T(\nu, \hat{n}) Y_{lm}^*(\hat{n}), \quad (11)$$

where  $Y_{lm}(\theta, \phi)$  are the spherical harmonic functions. The cross-power spectra between different frequency bands can then be related to the spectrum of perturbations of the matter density:

$$\begin{aligned} C_l(\nu_1, \nu_2) &\equiv \langle a_{lm}(\nu_1) a_{lm}^*(\nu_2) \rangle \\ &= \frac{2}{\pi} \int_0^\infty dk k^2 P_0(k) \omega_{l,1}(k) \omega_{l,2}(k). \end{aligned} \quad (12)$$

Here,  $P_0(k)$  is the  $z=0$  matter power spectrum, and we have defined the window functions

$$\omega_{l,i}(k) \equiv \int_0^\infty dz \phi_i(z) D(z) [b(z) j_l(k\chi) - f(z) j_l''(k\chi)], \quad (13)$$

where  $j_l$  is the  $l$ th spherical Bessel function,  $\phi_i$  is the selection function for the  $i$ th frequency band (given in terms of redshift),  $D(z)$  is the growth factor and  $f(z) = d \ln D / d \ln a$  is the growth rate. Here, we have used the Kaiser approximation (Kaiser 1987) for RSDs and have assumed that the H I density is linearly biased with respect to the matter density ( $\delta_{H\text{I}} = b(z) \delta_M$ ).

For the results shown here, the selection functions  $\phi_i(z)$  are just top-hat windows in the redshift range covered by a given frequency band. The effect of an instrumental beam or a finite pixel size can be included in the  $C_l$ ’s by scaling them with the corresponding angular window function  $C_l \rightarrow |b_l|^2 C_l$ , where  $b_l$  are the Legendre coefficients of the real space window  $b(\theta)$ :

$$b_l = 2\pi \int_{-1}^1 L_l(x) b(x \equiv \cos \theta) dx. \quad (14)$$

For top-hat and Gaussian beams, for instance, the respective window functions are

$$b_l^{\text{TH}} = \frac{L_{l-1}(\cos \theta_{\text{TH}}) - L_{l+1}(\cos \theta_{\text{TH}})}{(2l+1)(1 - \cos \theta_{\text{TH}})}, \quad (15)$$

$$b_l^{\text{G}} = \exp\left(-\frac{l(l+1)\theta_{\text{G}}^2}{2}\right). \quad (16)$$

#### 2.3.2 Radial clustering

Formally, the only way to analyse the clustering pattern of the H I distribution along the line of sight (LOS) is through the angular cross-correlation of different frequency channels. The main reason for this limitation is the fact that, while the matter distribution is expected to be isotropic, it is not homogeneous in the radial direction when observed on the lightcone: the growth of structure as well as the bias and the H I fraction, for example, evolve with  $z$ .

In practice, however, it is possible to explore the radial distribution if one is willing to accept a few approximations. Within the flat-sky approximation and for an infinitesimally thin frequency bin, we can describe a given angular pixel as a circle of radius  $R \simeq r(z) \Delta\theta$ , where  $\Delta\theta$  is the pixel angular resolution and  $r(z)$  is the comoving distance to the redshift corresponding to the frequency  $\nu$ . Then, we can assume that the temperature anisotropy in that pixel corresponds to the average temperature within this circle:

$$\delta_T(\nu, \mathbf{n}) \equiv \frac{\Delta T(\nu, \mathbf{n})}{T(\nu)} = \int \int_{C_R} \frac{d\mathbf{s}_\perp}{\pi R^2} \delta_{H\text{I}}(r_\parallel, \mathbf{r}_\perp - \mathbf{s}_\perp). \quad (17)$$

Here we have made use of equation (10) and have defined the radial and angular coordinates  $r_{\parallel} \equiv \chi(z)$ ,  $\mathbf{r}_{\perp} = r(z)\mathbf{n}$ . Note that all integrals with respect to transverse ( $\perp$ ) vectors are two-dimensional.

Now, let us assume that we take a finite frequency bin that is thick enough to study the radial clustering over the relevant scales but narrow enough so that we can assume that no significant evolution takes place within the corresponding range of redshifts. We can then study the two-point statistics for pixels in the same angular location but corresponding to different frequencies within our bin. Let us first define the LOS Fourier transform of  $\delta_T$ , which can be related to the Fourier modes of the three-dimensional overdensity field  $\delta_{H\text{I}}$ :

$$\begin{aligned}\tilde{\delta}_{\parallel}(k_{\parallel}) &\equiv \int \frac{dr_{\parallel}}{\sqrt{2\pi}} \exp[i k_{\parallel} r_{\parallel}] \delta_T(r_{\parallel}) \\ &= \int \frac{d\mathbf{k}_{\perp}}{2\pi} \left( 2 \frac{J_1(k_{\perp} R)}{k_{\perp} R} \right) \tilde{\delta}_{H\text{I}}(k_{\parallel}, \mathbf{k}_{\perp}).\end{aligned}\quad (18)$$

Here  $J_1(x)$  is the first-order cylindrical Bessel function of the first kind and  $\tilde{\delta}_{H\text{I}}(k_{\parallel}, \mathbf{k}_{\perp})$  is the Fourier component of  $\delta_{H\text{I}}$  for the wavevector with LOS and transverse components  $k_{\parallel}$  and  $\mathbf{k}_{\perp}$ , respectively. We can then define the radial power spectrum

$$\langle \tilde{\delta}_{\parallel}(k_{\parallel}) \tilde{\delta}_{\parallel}^*(k'_{\parallel}) \rangle \equiv \delta^D(k_{\parallel} - k'_{\parallel}) P_{\parallel}(k_{\parallel}), \quad (19)$$

which is directly related to the three-dimensional spectrum of neutral hydrogen:

$$P_{\parallel}(k_{\parallel}) = \int_0^{\infty} \frac{dk_{\perp} k_{\perp}}{2\pi} \left[ 2 \frac{J_1(k_{\perp} R)}{k_{\perp} R} \right]^2 P_{H\text{I}}(k_{\parallel}, k_{\perp}). \quad (20)$$

For the results shown in Section 5.1, we have used a simplified model for  $P_{H\text{I}}$ , neglecting non-linearities and including RSDs through the Kaiser approximation:

$$P_{H\text{I}}(k_{\parallel}, k_{\perp}) = D^2(z) \left( b(z) + f(z) \frac{k_{\parallel}^2}{k^2} \right)^2 P_0(k). \quad (21)$$

Note also that the effective finite width of the frequency shells used to calculate the power spectrum can be taken into account by scaling equation (20) with the square of the window function

$$W(k_{\parallel} \delta\chi) = 2 \frac{\sin(k_{\parallel} \delta\chi/2)}{k_{\parallel} \delta\chi}, \quad (22)$$

where  $\delta\chi = \frac{c(1+z)^2}{v_{21} H(z)} \delta\nu$  is the comoving scale corresponding to the shell width  $\delta\nu$ .

### 3 SIMULATING THE COSMOLOGICAL SIGNAL

Putting together the results described in Sections 2.1 and 2.2, the method used by our code to generate fast mock H I catalogues is as follows.

(i) Consider a cubic box of comoving size  $L$  and divide it into  $N_{\text{grid}}^3$  cubical cells of size  $l_c \equiv L/N_{\text{grid}}$ . This will determine the scales probed by the catalogue:  $2\pi/L \lesssim k \lesssim 2\pi/l_c$ .

(ii) Generate a realization of the Fourier-space Gaussian overdensity field at  $z = 0$  by producing Gaussian random numbers with variance

$$\sigma^2(k) \equiv \left( \frac{L}{2\pi} \right)^3 P_0(k). \quad (23)$$

This is done in a Fourier-space grid with  $\mathbf{k} = \mathbf{n} 2\pi/L$  with  $-N_{\text{grid}}/2 \leq n_i \leq N_{\text{grid}}/2$ .

At the same time, the  $z = 0$  velocity potential can be calculated from the overdensity field as

$$\varphi_k(z = 0) = f_0 H_0 \frac{\delta_k(z = 0)}{k^2}. \quad (24)$$

(iii) Transform these fields to configuration space using a fast Fourier transform, and calculate the radial velocity at each cell by projecting the gradient of the velocity potential along the LOS. The direction of the LOS will depend on the position of the observer inside the box (since we want to produce full-sky catalogues, we take this to be the centre of the box). This will yield the Gaussian overdensity  $\delta_G$  and radial velocity  $v_r$  fields at  $z = 0$ .

(iv) Calculate the overdensity field and radial velocity in the light-cone by computing the redshift to each cell through the distance–redshift relation (equation 9), and evolving the fields self-similarly to that redshift.

At the same time, we may perform the lognormal transformation on the Gaussian overdensity field. Thus, in a cell at  $\mathbf{x}$  with redshift  $z(\mathbf{x})$ , the overdensity and radial velocity are given by

$$1 + \delta_{H\text{I}}(\mathbf{x}) = \exp[G(z)\delta_G(\mathbf{x}, z = 0) - G^2(z)\sigma_G^2/2], \quad (25)$$

$$v_r(\mathbf{x}) = \frac{f(z)H(z)D(z)}{(1+z)f_0 H_0} v_r(\mathbf{x}, z = 0), \quad (26)$$

where  $\sigma_G^2 \equiv \langle \delta_G^2 \rangle$  is the variance of the Gaussian overdensity at  $z = 0$  and the factor  $G(z) \equiv D(z)b(z)$  accounts both for the growth of perturbations and for a possible linear galaxy bias  $b$ .

(v) We calculate the total H I mass stored in each cell through

$$M_{H\text{I}} = (2.775 \times 10^{11} \text{ M}_{\odot}) \left( \frac{l_c}{\text{Mpc } h^{-1}} \right)^3 \frac{\Omega_b x_{H\text{I}}(z)}{h} (1 + \delta_{H\text{I}}).$$

(vi) We divide the box into spherical shells, which are in turn pixelized to yield maps of the 21 cm brightness temperature at different frequency bands (corresponding to the width of the different radial shells). To each pixel, we assign the temperature associated with the hydrogen mass enclosed within it (equation 8). This implies interpolating between the Cartesian mass grid and the spherical pixels. We do this interpolation through a Monte Carlo integration: a number  $N \sim 10$  of points are randomly placed inside each cell, and a mass of  $M_{H\text{I}}/N$  is assigned to each of them. We then assign each point to a spherical shell and angular pixel according to its redshift and angular position. At this stage, we implement RSDs by perturbing the cosmological redshift of each point with the redshift distortion corresponding to the radial velocity in the cell  $\Delta z_{\text{RSD}} = c(1+z)v_r(\mathbf{x})$ . Our code uses the HEALPIX pixelization scheme (Górski et al. 2005).

As can be seen, the above recipe relies on two extra ingredients (besides the underlying cosmological model): we need a model for the evolution of the H I fraction  $x_{H\text{I}}$  and its bias with respect to the matter density  $b(z)$ .

### 4 FOREGROUNDS

Probably the main challenge of 21 cm observations for cosmology is the presence of galactic and extragalactic foregrounds with amplitudes several orders of magnitude larger than the cosmological signal – a situation which in many ways mimics that of the analysis of the CMB. However, the spectral smoothness of the foregrounds (or at least their clearly identifiable frequency dependence) should make it possible to subtract them.

Five different types of foregrounds have been implemented in the present version of the code: unpolarized and polarized galactic synchrotron, galactic and extragalactic free–free emission and emission



**Table 1.** Foreground  $C_l(v_1, v_2)$  model from Santos et al. (2005) for the pivot values  $l_{\text{ref}} = 1000$  and  $v_{\text{ref}} = 130$  MHz.

Foreground	A (mK <sup>2</sup> )	$\beta$	$\alpha$	$\xi$
Galactic synchrotron	700	2.4	2.80	4.0
Point sources	57	1.1	2.07	1.0
Galactic free-free	0.088	3.0	2.15	35
Extragalactic free-free	0.014	1.0	2.10	35

from extragalactic point radio sources. We have classified these into two categories, isotropic and anisotropic, according to their angular distribution on the sky, and different methods were used to simulate each of them. These methods are based on those used by other groups to simulate radio foregrounds (Jelić et al. 2010; Shaw et al. 2013, 2014).

We must note that the use of the term ‘foregrounds’ when referring to these contaminants for intensity mapping can be a misnomer: some of the sources of these contaminants are in fact ‘behind’ part of the H I signal (e.g. point sources can be found at very high redshifts). The use of this term, however, has become traditional and is widespread in the literature, and therefore we will use it here, trusting that it will not lead to confusion.

#### 4.1 Isotropic foregrounds

Assuming that the foregrounds are Gaussianly distributed, any statistically isotropic emission can be entirely modelled in terms of the frequency-space angular power spectra:

$$\langle a_{lm}(v_1) a_{l'm'}^*(v_2) \rangle = \delta_{ll'} \delta_{mm'} C_l(v_1, v_2). \quad (27)$$

We have simulated the emission from extragalactic point sources and free-free foregrounds according to this model.<sup>2</sup>

We have followed Santos, Cooray & Knox (2005, SCK from here on) in modelling the power spectrum as

$$C_l(v_1, v_2) = A \left( \frac{l_{\text{ref}}}{l} \right)^\beta \left( \frac{v_{\text{ref}}^2}{v_1 v_2} \right)^\alpha \exp \left( -\frac{\log^2(v_1/v_2)}{2\xi^2} \right), \quad (28)$$

where  $\xi$  is the frequency-space correlation length of the emission, which regulates its spectral smoothness (foregrounds with smaller  $\xi$  will be less smooth in frequency, and will be therefore more challenging to subtract). The parameters used for the different foregrounds were taken from SCK and are listed in Table 1.

In order to include the frequency decorrelation, we have

- (i) diagonalized the matrix  $C_{ij} \equiv C_l(v_i, v_j)/[A (l_{\text{ref}}/l)^\beta]$ ;
- (ii) generated independent random Gaussian realizations in the diagonal basis with power spectrum

$$C_l^n = A \left( \frac{l_{\text{ref}}}{l} \right)^\beta \lambda^n, \quad (29)$$

where  $\lambda^n$  is the  $n$ th eigenvalue of  $C_{ij}$ ; and

- (iii) rotated these realizations back to frequency space.

##### 4.1.1 Note on point sources

Extragalactic point sources are objects beyond our galaxy emitting in radio (e.g. AGNs). These objects should be clustered following

<sup>2</sup> Obviously galactic free-free emission is not homogeneous. However, due to its exceptionally smooth frequency dependence and its subdominant amplitude, we do not believe that a more sophisticated modelling is required at this stage.

the same matter distribution as the cosmological signal. Hence, this foreground is potentially correlated with the cosmological signal if the redshift distribution of point sources overlaps with the redshift domain for which we aim to measure  $\delta_{\text{HI}}$  (which could be the case for intensity mapping). Therefore, in order to take this potential correlation into account, it would be desirable to be able to generate these point sources directly using the same matter distribution as was used for the cosmological signal. We have implemented this in our code keeping two main studies in mind.

First, we would like to test the validity of approximating the emission from point sources as Gaussian realizations of the SCK model. To do so, we compared this approach with a more realistic one: we Poisson-sampled the matter density field used for the cosmological signal with point sources following the luminosity function corresponding to star-forming galaxies that was used in Wilman et al. (2008). Each point source was given a luminosity and a random spectral index  $\alpha$  following a normal distribution

$$P(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} e^{-(\alpha-\alpha_0)^2/2\sigma_\alpha^2}, \quad (30)$$

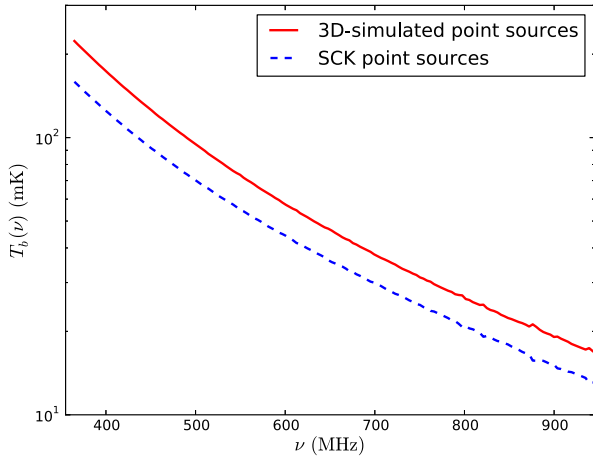
with  $\alpha_0 = 0.07$  (Wang et al. 2006). We verified that, even for large values of  $\sigma_\alpha \sim 5$ –10, the combined emission from point sources in the same angular pixel is extremely smooth and that in fact the frequency correlation length  $\xi = 1$  used above is a conservative lower bound.

Secondly, since point sources are correlated with the cosmological signal, subtracting the former could potentially destroy (or degrade) the latter, an effect that our simulation, based on the SCK model, would not capture. We have attempted a small study of this effect here, although a more elaborate analysis is necessary to thoroughly address this potential issue.

For a simulation of the cosmological signal (corresponding to the FAST parameters in Table 2), we generated two simulations of the point-source emission, one based on the SCK model and another one produced by sampling the same density field and giving each point source a luminosity and spectral energy distribution (SED) as was done in Wilman et al. (2008) for star-forming galaxies. We then produced two simplified observed sky maps by adding the cosmological signal and the two point-source simulations. For these simulations, we also imposed a Gaussian beam with  $\theta_{\text{FWHM}} = 0''.3$  and uncorrelated Gaussian noise with  $\sigma_T = 0.05$  mK. Fig. 1 shows the brightness temperature in a single pixel as a function of frequency for both simulations.

**Table 2.** Characteristics of the two types of simulations that were run for this work. The computational times correspond to real times using an 80-core shared-memory machine for the complete simulation (cosmological signal + foregrounds). The output sizes correspond to one single field (e.g. the cosmological signal).

	ACCURATE	FAST
$N_{\text{grid}}$	3072	2048
$L_{\text{box}}$ (Mpc $h^{-1}$ )	8150	8900
$\nu$ interval (MHz)	(405, 945)	(355, 945)
$N_\nu$	770	150
$z$ interval	(0.5, 2.5)	(0.5, 3)
$\delta r_{\parallel}$ (Mpc $h^{-1}$ )	$\sim 3.5$	$\sim 20$
$n_{\text{side}}$	512	512
Angular resolution (arcmin)	6.7	6.7
Computational time	$\sim 10$ h	$\sim 30$ min
Output size (GB, 1 field)	18	3.5



**Figure 1.** Brightness temperature as a function of frequency for a single pixel of two of our simulations containing the cosmological signal and two different realizations of point-source foregrounds. The red solid line shows the result for point sources simulated by sampling the density distribution that was used to generate the cosmological signal, while the blue dashed line corresponds to a random realization of the **SCK** model.

In order to study the potential loss of signal due to the subtraction of a foreground that is correlated with the signal, we cleaned the point-source foregrounds from these two simulations using the singular value decomposition method that was used in Switzer et al. (2013). For the **SCK** point-source maps, the cleaning method converged after subtracting six foreground modes, while only two modes were necessary for the 3D-simulated point sources. This is due to the fact that the **SCK** model assumes a very conservative frequency correlation length  $\xi$  for point sources, while for the 3D simulation, every point source has the same SED. We then calculated the residual maps  $T_{\text{clean}}(\nu) - T_{\text{signal}}(\nu)$  and computed their angular power spectra. We can measure the amount of signal loss by comparing these power spectra with the expected statistical errors in the cosmological signal:

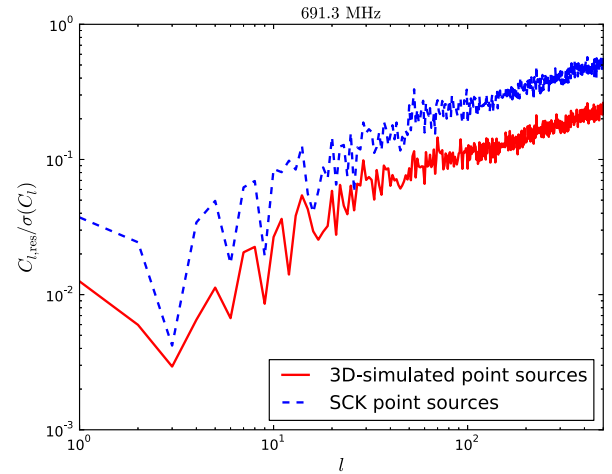
$$\sigma(C_l) \simeq \sqrt{\frac{2}{2l+1}} \left( C_l + \frac{1}{\sigma_T^2} \right). \quad (31)$$

We have plotted the ratio of these two quantities in Fig. 2. As is shown, the residual  $C_l$  is well below statistical uncertainties on all scales for both simulations, but a larger fraction of the signal is lost for the **SCK** foregrounds. This shows that the smoothness in frequency of the foregrounds is in fact more relevant than the potential correlation of point sources with the cosmological signal.

The publicly available version of the code allows the user to simulate one population of point sources following a particular luminosity function, large-scale bias and SED. However, we discourage the use of this feature, since the enormous number of point sources increases the computational time significantly and, as has been shown here, this method has no practical advantage over the empirical **SCK** model as far as foreground-cleaning studies are concerned.

#### 4.1.2 Line foregrounds

Another possible contaminant for H I intensity mapping could be due to the line emission of other sources (either galactic or extragalactic), and since this component is not spectrally smooth, it might be impossible to subtract. However, one of the main advantages of the 21 cm line is that it is very isolated in frequency, and therefore any source emitting in most of the potentially harmful



**Figure 2.** Ratio of the power spectrum of the foreground cleaning residuals to the statistical errors in the cosmological power spectrum for a frequency bin at  $\bar{\nu} = 691.3$  MHz. The red solid line shows the result for maps containing point sources sampled from the density distribution, while the blue dashed line corresponds to the **SCK** model for point sources.

lines would have to be at a very high redshift for it to contaminate the H I signal. Only a few molecular lines, such as the hydroxyl radical (OH,  $\sim 1600$ – $1700$  MHz), lie close to 21 cm, and their signal is known to be very weak (Gong et al. 2011). Therefore, unlike other lines, intensity mapping for H I should be virtually free of line confusion, and we have not attempted to include this contribution in our simulations.

## 4.2 Galactic synchrotron

As can be seen in Table 1, galactic synchrotron is by far the largest contribution to the total emission in the frequencies of interest. For this reason, it is very important to model it as realistically as possible.

One of the key differences between galactic and extragalactic foregrounds is their angular structure: the intensity grows steeply towards the galactic plane, and everywhere the angular spectral tilt of galactic foregrounds is redder, making large angular structures more relevant. Subtracting a foreground with such a large structure could potentially erase or distort the cosmological signal on the largest angular scales, which could be a problem in order to measure large-scale observables, such as non-Gaussianity in the two-point clustering (Camera et al. 2013).

Furthermore, linearly polarized radiation changes its polarization angle in as it traverses the galaxy due to Faraday rotation, an effect that is frequency dependent and therefore not spectrally smooth. Hence, if part of the polarized foreground is leaked into the unpolarized part due to instrumental issues, it could be extremely difficult to subtract (e.g. De & Tashiro 2014).

In this section, we will first briefly describe the basics of galactic synchrotron emission and then explain the methods used in the simulations. These methods are largely based on those used by Shaw et al. (2013, 2014).

### 4.2.1 Theoretical background

A detailed description of the principles of galactic synchrotron emission can be found, for example, in Pacholczyk (1970) and Rybicki & Lightman (1986). The galactic synchrotron radiation is caused

by cosmic ray (CR) electrons interacting with the galactic magnetic field. The intensity of the emission coming from a volume element  $dV = s^2 ds \delta\Omega$  in a frequency interval  $\delta\nu$  is given by the emission coefficient  $j(s, \hat{n}, \nu)$ . For the total and polarized synchrotron radiation, this is given by

$$j_I(s, \hat{n}, \nu) = C_I \left( \frac{2\pi m_e c}{3e} \nu \right)^{\frac{1-p}{2}} n_{\text{CR}} B_{\perp}^{\frac{p+1}{2}},$$

$$j_P(s, \hat{n}, \nu) = C_P \left( \frac{2\pi m_e c}{3e} \nu \right)^{\frac{1-p}{2}} n_{\text{CR}} B_{\perp}^{\frac{p+1}{2}} e^{i2\phi_0}, \quad (32)$$

where  $n_{\text{CR}}$  is the cosmic ray electron density,  $B_{\perp}$  is the transverse galactic magnetic field,  $\phi_0$  is the initial polarization angle and we are assuming a power-law energy distribution for the CR electrons  $N(E) \propto E^{-p}$ . The coefficients  $C_{I,P}$  are given by

$$C_I = \frac{\sqrt{3}e^3}{4\pi m_e c^2(p+1)} \Gamma\left(\frac{3p-1}{12}\right) \Gamma\left(\frac{3p+19}{12}\right),$$

$$C_P = \Pi_0 C_I, \quad \Pi_0 = \frac{3p+3}{3p+7}. \quad (33)$$

Assuming a spectral index  $p \sim 2.5$ , the intrinsic polarized fraction is  $\Pi_0 \sim 0.7$ . Note that we have grouped the  $Q$  and  $U$  Stokes parameters into one single complex number labelled by  $P$ . The total intensity measured from the Earth is the LOS integral of these emissivities.

As the synchrotron photons traverse the magnetized interstellar medium, their polarization angle changes, an effect known as *Faraday rotation* [see Rybicki & Lightman (1986) for further details]. The observed polarization angle is thus not the same as the original one at the point of emission  $\phi_0$ . Both angles are related by  $\phi = \phi_0 + \psi(s, \hat{n})\lambda^2$ , where  $\lambda = c/\nu$  and  $\psi$  is the so-called Faraday rotation measure, given in terms of the longitudinal galactic magnetic field and the free electron density as

$$\psi(s, \hat{n}) = \frac{e^3}{2\pi(m_e c^2)^2} \int_0^s ds' n_e(s', \hat{n}) B_{\parallel}(s', \hat{n}). \quad (34)$$

Thus, the polarized synchrotron intensity can be written as

$$I_P(\nu, \hat{n}) = \Pi_0 \int_0^\infty ds j_I(s, \hat{n}, \nu) e^{2i\phi_0(s, \hat{n})} e^{i\psi(s, \hat{n})x_\nu}, \quad (35)$$

where  $x_\nu = 2(c/\nu)^2$ .

#### 4.2.2 Unpolarized emission

The unpolarized synchrotron radiation should be spectrally smooth and, far from the galactic plane, we should be able to model it as an isotropic field, just like we did in Section 4.1. However, we would also like to include the shape of the emission from the galactic plane, for which we have used the following recipe.

(i) The Haslam map (Haslam et al. 1982)<sup>3</sup> contains the full-sky synchrotron emission at  $\nu_H = 408$  MHz. This map is further filtered on angular scales  $l \gtrsim 200$  corresponding to its angular resolution ( $\theta \sim 1^\circ$ ). As a first approximation, we could simulate the emission in other frequencies by extrapolating the Haslam map with some spectral index. We have used the Planck Sky Model (PSM; Delabrouille et al. 2013) to generate a full-sky map of the synchrotron

spectral index  $\beta(\hat{n})$ . With these two maps, we calculate

$$T_0(\nu, \hat{n}) = T_{\text{Haslam}}(\hat{n}) \left( \frac{\nu_H}{\nu} \right)^{\beta(\hat{n})}. \quad (36)$$

(ii) Due to the poor resolution of the Haslam map, it would be desirable to add structure on smaller scales as well as to introduce decorrelation between different frequencies. We do this by generating a Gaussian realization of the  $C_l$  model used in Section 4.1, with the parameters listed in Table 1. These additional fluctuations should be constrained to yield 0 on the scales already given by the Haslam map at  $\nu = \nu_H$ . This can be guaranteed by constraining the eigenmode corresponding to the largest eigenvalue of the matrix  $C_{ij}$ .

Let us clarify this. The temperature fluctuation on the  $i$ th frequency band is given by

$$\delta T(\nu_i) = \sum_{n=1} B_i^n \delta \tilde{T}_n, \quad (37)$$

where  $\hat{B}$  is the orthogonal matrix that diagonalizes  $C_{ij}$  and  $\delta \tilde{T}_n$  is the uncorrelated realization of the diagonal  $C_l$ 's corresponding to the  $n$ th eigenvalue of  $C_{ij}$ . We want  $\delta T(\nu_H) = 0$  on the scales constrained by the Haslam map. Assuming that the highest eigenvalue corresponds to  $n = 1$ , we constrain the corresponding eigenmode on these scales to be

$$\delta \tilde{T}_1 = - \sum_{n=2} \frac{B_{i_H}^n}{B_{i_H}^1} \delta \tilde{T}_n, \quad (38)$$

where  $i_H$  is the frequency bin corresponding to  $\nu_H$ .

(iii) The frequency-decorrelated unpolarized synchrotron emission including the galaxy is then given by  $T = T_0 + \delta T$ .

#### 4.2.3 Polarized foregrounds

The frequency-dependent Faraday rotation affecting polarized synchrotron emission makes any leakage of this foreground into the unpolarized signal a potentially dangerous contribution. Thus, in order to assess the feasibility of this subtraction or, alternatively, the level of polarization leakage that can be allowed in any intensity mapping survey, it is necessary to have a correct description of this foreground.

The existing observational data regarding polarized synchrotron emission are limited to isolated radio and microwave frequency bands (Wolleben et al. 2006; Testori, Reich & Reich 2008), and the structure of the galactic magnetic field is poorly understood (e.g. Haverkorn et al. 2008). This situation will improve in the future (e.g. Wolleben et al. 2009), but for the moment it is difficult to develop a reliable model of the polarized foregrounds in radio frequencies based only on observations. Alternatively, one can use existing models of the galactic magnetic field and CR and thermal electron densities in order to obtain more realistic predictions. This is the approach taken in HAMMURABI (Waelkens et al. 2009), a computer code that generates a 3D simulation of the Milky Way based on different models and then performs the LOS integration to generate temperature maps at different frequencies. We have instead followed a different approach, based on the statistical properties of the synchrotron emission in the space of Faraday depths. As described in Section 5.2.1, we have used HAMMURABI a posteriori to validate this model.

For each LOS  $\hat{n}$ , we can use the Faraday depth  $\psi(s, \hat{n})$  as LOS coordinate, instead of  $s$ , and rewrite equation (35) as

$$I_P(\nu, \hat{n}) = \int d\psi k(\psi, \hat{n}, \nu) e^{i\psi x_\nu}, \quad (39)$$

<sup>3</sup> We used the filtered and de-sourced map available at NASA's Legacy Archive for Microwave Background Data Analysis – <http://lambda.gsfc.nasa.gov>

where  $k(\psi_0) = \int ds \delta(\psi(s) - \psi_0) j_1(s) e^{2i\phi_0(s)}$  is the collective emission from regions with Faraday depth  $\psi$ . As a first approximation, we can assume that the spectral dependence of the emission is basically the same at all depths and can be factorized:  $k(\psi, \hat{n}, \nu) = b(\nu, \hat{n}) k_0(\psi, \hat{n})$ .

Oppermann et al. (2012) used extragalactic point sources to measure the Faraday depth to the end of the Milky Way  $\psi_\infty(\hat{n})$ . We can use their map to inspect the distribution of  $\psi$  in different directions (see the bottom-left panel in Fig. 6). Since in this map  $\psi$  seems to take equally negative and positive values, we could model  $\psi$  as being normally distributed around 0 with some variance  $\sigma^2(\hat{n})$ . We can estimate this variance from the maps of  $\psi_\infty$  by smoothing  $\psi_\infty^2$  on a large angular scale (we have used  $5^\circ$ ). Then, assuming that the collective emission at some  $\psi$  is just proportional to the number of regions with that Faraday depth, we can model  $k_0$  as

$$k_0(\psi, \hat{n}) = B \exp \left[ -\frac{1}{2} \left( \frac{\psi}{\sigma(\hat{n})} \right)^2 \right] \mu(\psi, \hat{n}). \quad (40)$$

Lacking a better motivated model, we can assume that the field  $\mu(\psi, \hat{n})$  has the same angular structure as the unpolarized emission, and that it is correlated in Faraday space on scales smaller than some correlation length  $\xi_\psi$

$$\langle \mu_{lm}(\psi) \mu_{l'm'}^*(\psi') \rangle \propto \delta_{ll'} \delta_{mm'} \left( \frac{l_{\text{ref}}}{l} \right)^\beta e^{-\frac{1}{2} \left[ \frac{\psi - \psi'}{\xi_\psi} \right]^2}, \quad (41)$$

where the choice of a proportionality constant is degenerate with  $B$ . In order to diagonalize this covariance, we can define the Fourier transform of  $\mu(\psi)$

$$\tilde{\mu}(x) \equiv \int \frac{d\psi}{\sqrt{2\pi}} \mu(\psi) e^{i\psi x}, \quad (42)$$

which is uncorrelated for different values of  $x$  with variance  $\propto l^{-\beta} e^{-(\xi_\psi x)^2/2}$ . It is then trivial to generate Gaussian realizations of the  $\tilde{\mu}(x)$ 's.

In terms of these functions, the integral in equation (39) can be written as

$$I_P(\nu, \hat{n}) = B' b(\nu, \hat{n}) \int_{-\infty}^{\infty} dx \tilde{\mu}(x, \hat{n}) e^{-\frac{(x - x_\nu)^2}{2\sigma^2(\hat{n})}}. \quad (43)$$

Modelling the frequency dependence as  $b(\nu, \hat{n}) = I_1(\nu, \hat{n})/I_1(\nu_0, \hat{n})$ , we are left with two free parameters: the overall amplitude  $B'$  and the correlation length in Faraday space  $\xi_\psi$ . There are two observational facts that we would like our simulated catalogues to mimic.

(i) First, the synchrotron emission is more strongly depolarized closer to the galactic plane. This is a sensible result: the magnetic field is stronger in that direction, and there are a larger number of emitting regions adding up incoherently. The ‘width’ of a given LOS in Faraday space is described in our simulation by the value of  $\sigma(\hat{n})$ , which we obtained from the map of Oppermann et al. (2012). Therefore, this effect should be automatically included.

(ii) At high frequencies ( $x_\nu \rightarrow 0$ ), the average depolarization at high galactic latitudes is  $\sqrt{|I_P|^2/I_I} \sim 0.2\text{--}0.3$ . We can use this observation to fix the amplitude  $B'$ .

We are left with one free parameter, the correlation length in Faraday space  $\xi_\psi$ , which governs the frequency structure of the polarized emission. As described below (Section 5.2.1), we have determined the value of this parameter by comparing the frequency dependence of our simulated maps to the one obtained using the more complex and realistic methods of HAMMURABI.

To summarize, the method we have used to simulate the polarized synchrotron foreground is as follows.

(i) Generate a Gaussian random realization of the  $\mu(x, \hat{n})$  with power spectrum  $C_l \propto l^{-\beta} e^{-x^2 \xi_\psi^2/2}$  (arbitrary normalization). The range and resolution in  $x$  for these realizations will be governed by the convergence of the integral in equation (43) for all values of  $\sigma(\hat{n})$ .

(ii) For each LOS  $\hat{n}$  and frequency  $\nu$ , we calculate the integral in equation (43) by summing over the realizations of  $\mu$ . Note that each LOS has its own ‘Faraday width’  $\sigma(\hat{n})$ , which will determine the relative depolarization of that LOS.

(iii) Fix the proportionality constant in equation (43) by requiring the average polarized fraction at large (microwave) frequencies and high galactic latitudes to be  $\sim 0.2\text{--}0.3$ , as measured by CMB experiments (Kogut et al. 2007).

We are mainly interested in the amount of polarized intensity that is leaked into the unpolarized part due to instrumental issues. The default version of our code implements this simply as a constant fraction of the Stokes parameter  $Q$ :

$$T_{\text{leak}}(\nu, \hat{n}) = \epsilon_p T_{\text{syn}}^Q(\nu, \hat{n}). \quad (44)$$

We would like to emphasize that this model is overly simple and not at all realistic, since the leakage is defined with respect to telescope coordinates, and thus will correspond to different combinations of  $Q$  and  $U$  depending on the time of observation. Our aim here is to provide only an order-of-magnitude comparison between the polarization leakage and the cosmological signal (e.g. Fig. 8), but for any practical application a more precise parametrization would be necessary, using the full  $Q$  and  $U$  maps provided by the code.

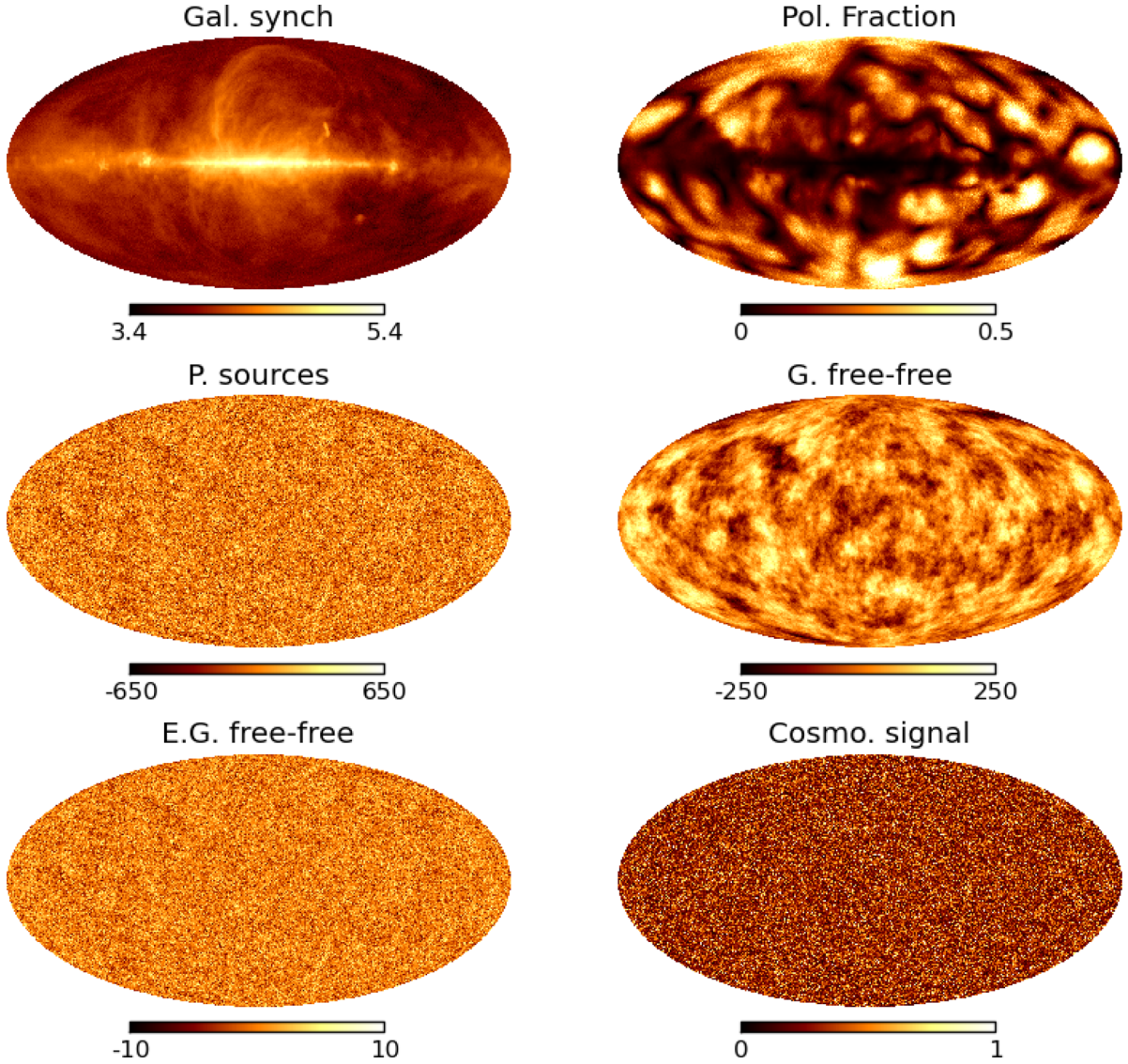
## 5 CODE VALIDATION

In order to validate the method, we have run several full simulations including the cosmological signal and foregrounds, and explored the qualitative behaviour of the different components. Fig. 3 shows the full-sky maps generated for the different fields in a frequency bin  $\nu \in (563, 567)$  MHz.

We have generated two different kinds of simulations with different spatial resolutions. The parameters for these are given in Table 2. The simulations labelled ACCURATE use a good spatial resolution ( $l_c \simeq 2.7 \text{ Mpc } h^{-1}$ ) and narrow frequency bins ( $\delta\nu \simeq 0.7 \text{ MHz}$ ). They were used to compare their radial and angular power spectra with the theoretical predictions described in Section 2.3, as well as to evaluate the validity of our model for the polarized synchrotron foregrounds. The simulations labelled FAST were run with a coarser grid ( $l_c \simeq 4.3 \text{ Mpc } h^{-1}$ ) and wider frequency bins ( $\delta\nu \simeq 4 \text{ MHz}$ ). They represent a lower quality but faster version of the ACCURATE simulations and were used mainly to test the maximum speed that could be attained for a full valid simulation.

We have not attempted to include any realistic instrumental effects in our simulations besides assuming a 1 per cent polarization leakage, since we would like the output to be applicable to any kind of experiment after a suitable post-processing. However, in order to show the lack of sensitivity of intensity mapping to small angular scales, we have implemented an angular Gaussian beam of width  $\theta_{\text{FWHM}} = 0.3^\circ$ . At an intermediate redshift of 1.75, this corresponds to a comoving scale of about  $7.6 \text{ Mpc } h^{-1}$  for the cosmological parameters below.





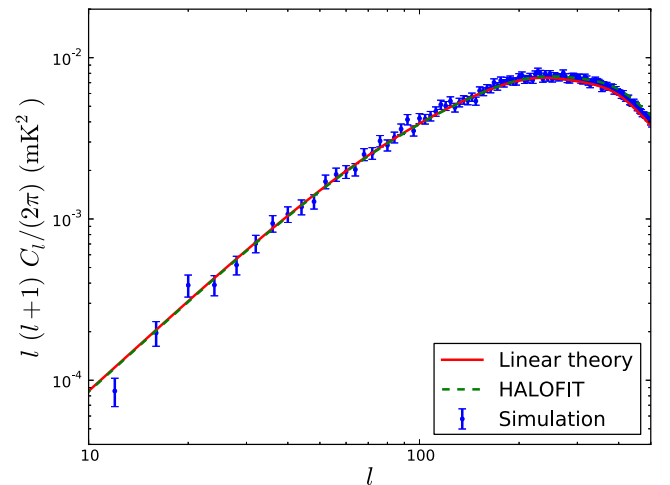
**Figure 3.** Full-sky maps of the different foregrounds and the cosmological signal for a frequency slice  $\nu \sim 565$  MHz. Temperatures are given in mK [with the top-left plot showing  $\log_{10}(T_{\text{synch}})$ ], except in the case of the polarized fraction (upper-right panel), which is dimensionless.

### 5.1 The cosmological signal

For the cosmological signal, we have used a  $\Lambda$  cold dark matter ( $\Lambda$ CDM) model with parameters  $(\Omega_M, \Omega_b, \Omega_k, h, w_0, w_a, \sigma_8, n_s) = (0.3, 0.049, 0, 0.67, -1, 0, 0.8, 0.96)$ , similar to the best-fitting values found by Planck Collaboration (2013b). The linear power spectrum used to generate the matter field was generated with CAMB (Lewis, Challinor & Lasenby 2000) using these parameters. As a model for the neutral fraction, we used  $x_{\text{HI}}(z) = 0.008(1+z)$ , which is a good fit to the two existing measurements at  $z = 0.01$  (Zwaan et al. 2005) and 1.5 (Wolfe, Gawiser & Prochaska 2005). For simplicity, we assumed that the H I is unbiased with  $b_{\text{HI}} = 1$ .

#### 5.1.1 Angular clustering

We have estimated the angular power spectrum of the cosmological signal for different frequency bins and compared it with the result predicted by linear theory. The comparison for the frequency bin at  $\bar{\nu} = 565$  MHz ( $z \sim 1.5$ ) can be seen in Fig. 4. The theoretical power spectrum was corrected for the finite pixel size using a



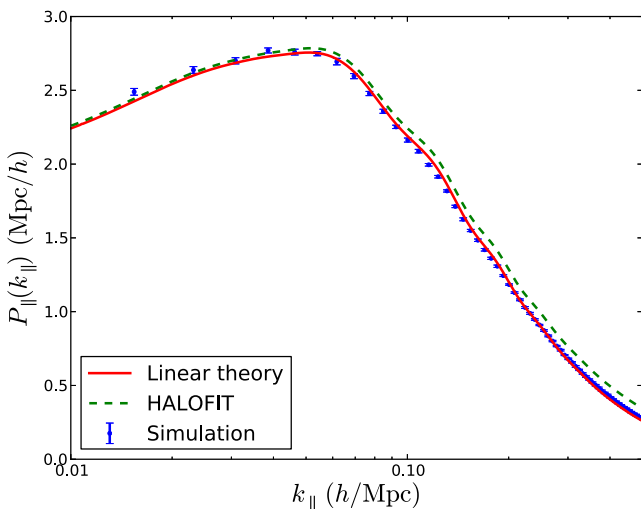
**Figure 4.** Angular power spectrum of the cosmological signal at  $\nu \sim 565$  MHz as measured in our simulation (blue dots with error bars) and according to linear theory (solid red line).

top-hat window function with a width equal to our HEALPIX pixel resolution ( $n_{\text{side}} = 512 \rightarrow \theta_{\text{TH}} \sim 0.06$ ). Rigorously speaking, this correction would only be valid if all the pixels had a circular shape; however, this correction is negligible compared with the window function due to the Gaussian beam, which removes most of the power above  $l \sim 400$ . Fig. 4 also shows an estimate of the non-linear angular power spectrum, using the  $P(k)$  predicted by HALOFIT (Smith et al. 2003); however, the difference with respect to the linear prediction is negligible due to the large angular beam.

### 5.1.2 Radial clustering

We have also verified that our simulations reproduce the correct clustering pattern along the LOS. We have done so by collecting all the frequency shells corresponding to the redshift interval  $1.4 < z < 2.1$ . For each pixel, this collection forms a vector with the values of the temperature fluctuation in different equispaced frequencies (which roughly correspond to equispaced radial distances). We perform a fast Fourier transform on each of these vectors to estimate  $\delta_{\parallel}(k_{\parallel})$  and estimate the radial power spectrum by averaging over all the pixels.

Fig. 5 shows the comparison between the thus measured radial power spectrum and the theoretical prediction outlined in Section 2.3 using a radial window function corresponding to the comoving width of the spherical shells in our ACCURATE simulations  $\delta\nu = 0.7 \text{ MHz} \rightarrow \delta r_{\parallel} \sim 4 \text{ Mpc } h^{-1}$ . As in the case of the angular power spectrum, the radial clustering is correctly reproduced by our simulations for a wide range of scales. The differences observed on large scales (small  $k_{\parallel}$ ) can be understood as being due to the fact that, while the frequency width is fixed for all the shells, the corresponding physical widths are frequency dependent and vary slightly ( $\sim 10$  per cent) across this frequency band, an effect that is not taken into account in our simplified model (Section 2.3.2). The non-linear prediction given by HALOFIT is also shown as a green line, which deviates significantly from the linear theory model on small scales (large  $k_{\parallel}$ ). We must caution the reader that this result is probably not quantitatively precise: the exact nature of different effects, such as RSDs or scale-dependent bias for H I, is not well understood on



**Figure 5.** Cosmological radial power spectrum for a bin  $1.4 < z < 2.1$ , corresponding to frequencies  $457 < \nu < 590 \text{ MHz}$ . The blue dots correspond to the measurement from an ACCURATE simulation, while the solid red line shows the linear theory prediction.

non-linear scales, and therefore our simplified model is probably not realistic in this regime.

## 5.2 Foregrounds

As explained in Section 4.1, the isotropic foregrounds were generated simply as random Gaussian realizations of the  $C_l$ 's in equation (28), using the parameters in Table 1. On the other hand, simulating the polarized and unpolarized synchrotron foregrounds requires the use of three additional external data sets: the 408 MHz Haslam map, the spectral index map from the PSM and the map of Faraday depths from Oppermann et al. (2012). These data sets are displayed in Fig. 6.

We have studied the angular distribution of these foregrounds as well as their frequency dependence. Fig. 7 shows the angular power spectra of the different foreground components compared to the cosmological signal that we expect to measure, which is several orders of magnitude smaller.

Fig. 8 shows the frequency dependence of the different components along LOS with different galactic latitudes ( $b = 70^\circ, 40^\circ$  and  $20^\circ$  from top to bottom). We can see that most foreground components are smooth in frequency, and should therefore be amenable to standard subtraction techniques. The leaked polarized synchrotron, on the other hand, has a non-trivial frequency structure, and could be extremely challenging to subtract. This problem becomes more important closer to the galactic plane, since the galaxy becomes ‘thicker’ in Faraday space, and the effects of Faraday rotation are more relevant. For this reason, special effort has been invested into verifying that our mock maps of the polarized synchrotron emission are statistically sensible.

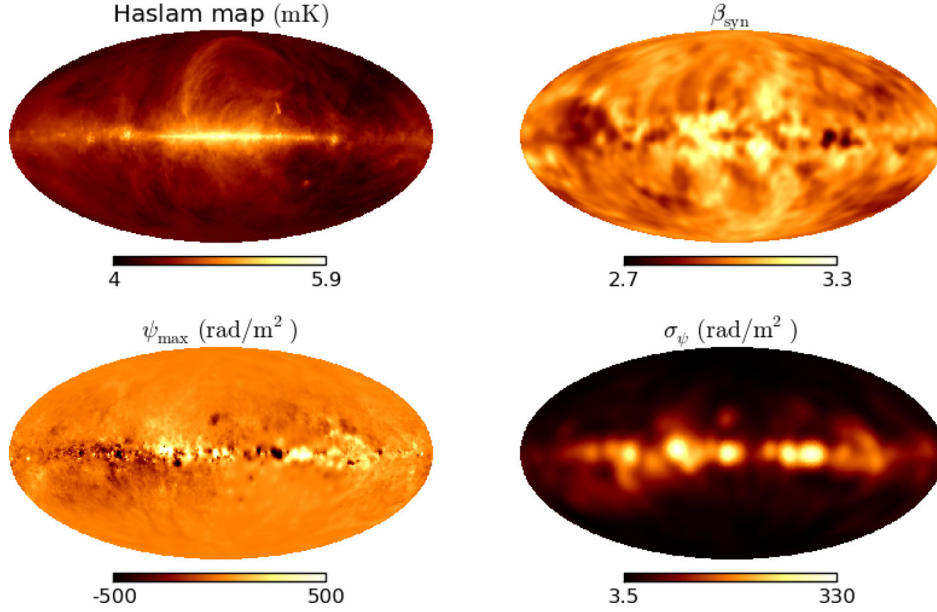
### 5.2.1 Polarized foregrounds

According to the model described in Section 4.2.3, the contribution of the polarized synchrotron to the foregrounds is described by two extra parameters, namely the Faraday-space correlation length  $\xi_{\psi}$  and the polarization leakage fraction  $\epsilon_p$ . While the latter depends entirely on the instrument design, the former has a direct effect on the degree of frequency decorrelation, which is key to determining whether or not it will be possible to subtract it.

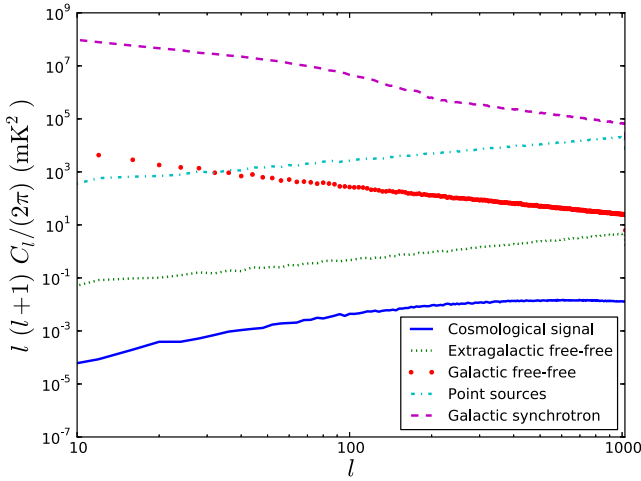
In order to validate our model for the polarized synchrotron foregrounds, we have run HAMMURABI using different parameters and spatial resolutions, and have compared the results with the maps generated by our code. In particular, we verified that the results shown below were qualitatively stable for the different models used by HAMMURABI to simulate the galactic magnetic field and CR electron density, the three-dimensional and angular resolution used by the code and the spectral tilt of the small-scale magnetic field. The fiducial simulation, for which we show results below, was run using the models for the magnetic field and CR density from Sun et al. (2008), three radial shells, an angular resolution parameter  $n_{\text{side}} = 256$  for the observation shell, a radial resolution of  $\sim 0.1 \text{ kpc}$  and a Cartesian grid with resolution  $\sim 0.07 \text{ kpc}$  [see Waelkens et al. (2009) for an overview of these parameters]. The rms variance of the small-scale magnetic field was set to  $3 \mu\text{G}$ . We generated synchrotron sky maps for 150 frequency bins between 945 and 355 MHz.

There are two main effects that we want our mock maps to reproduce.

- (i) The degree of frequency decorrelation produced by the frequency-dependent Faraday rotation is the main source of



**Figure 6.** External data sets used to simulate the unpolarized and polarized synchrotron foregrounds: the Haslam map (top left) (temperature given in mK), the synchrotron spectral index  $\beta(\hat{n})$  predicted by the PSM (top right), the map of Faraday depths  $\psi_\infty$  compiled by Oppermann et al. (2012, bottom left) and the Faraday widths  $\sigma(\hat{n})$  estimated by smoothing  $\psi_\infty^2$  (bottom right).



**Figure 7.** Angular power spectra of the different foreground components and the cosmological signal at  $\nu \sim 550$  MHz.

complications in terms of foreground subtraction. Thus, we must make sure that this decorrelation is correctly reproduced by our model. In order to quantify this decorrelation, we have computed the frequency-space power spectrum on different galactic latitudes for the HAMMURABI simulations and for our own mock data sets. This was done as follows.

- (a) We take the emission in  $Q$  for different frequencies for all the pixels in a wide interval of galactic latitudes (we used  $\Delta b = 10^\circ$ ).
- (b) We Fourier-transform the emission from each pixel in frequency space:

$$\tilde{Q}(k_\nu, \mathbf{n}) \propto \int d\nu Q(\nu, \mathbf{n}) e^{ik_\nu \nu}. \quad (45)$$

- (c) We compute the power spectrum  $P(k_\nu) \equiv \langle |\tilde{Q}(k_\nu)|^2 \rangle$  by averaging over all the pixels in the stripe.

As mentioned above, Faraday rotation should be more important towards the galactic centre. Hence, frequency decorrelation should increase and  $P(k_\nu)$  should become bluer at low  $b$ . We can use this effect to constrain the only free parameter of our model: the Faraday-space correlation length  $\xi_\psi$ . This effect is explicitly shown in Fig. 9, where we have plotted the normalized power spectra for three latitude stripes at  $b = 65^\circ, 40^\circ$  and  $20^\circ$ . We find that a value  $\xi_\psi \sim 0.5 \text{ rad m}^{-2}$  is able to qualitatively reproduce this effect.

- (ii) The polarized fraction

$$\Pi \equiv \left\langle \frac{\sqrt{Q^2 + U^2}}{I} \right\rangle \quad (46)$$

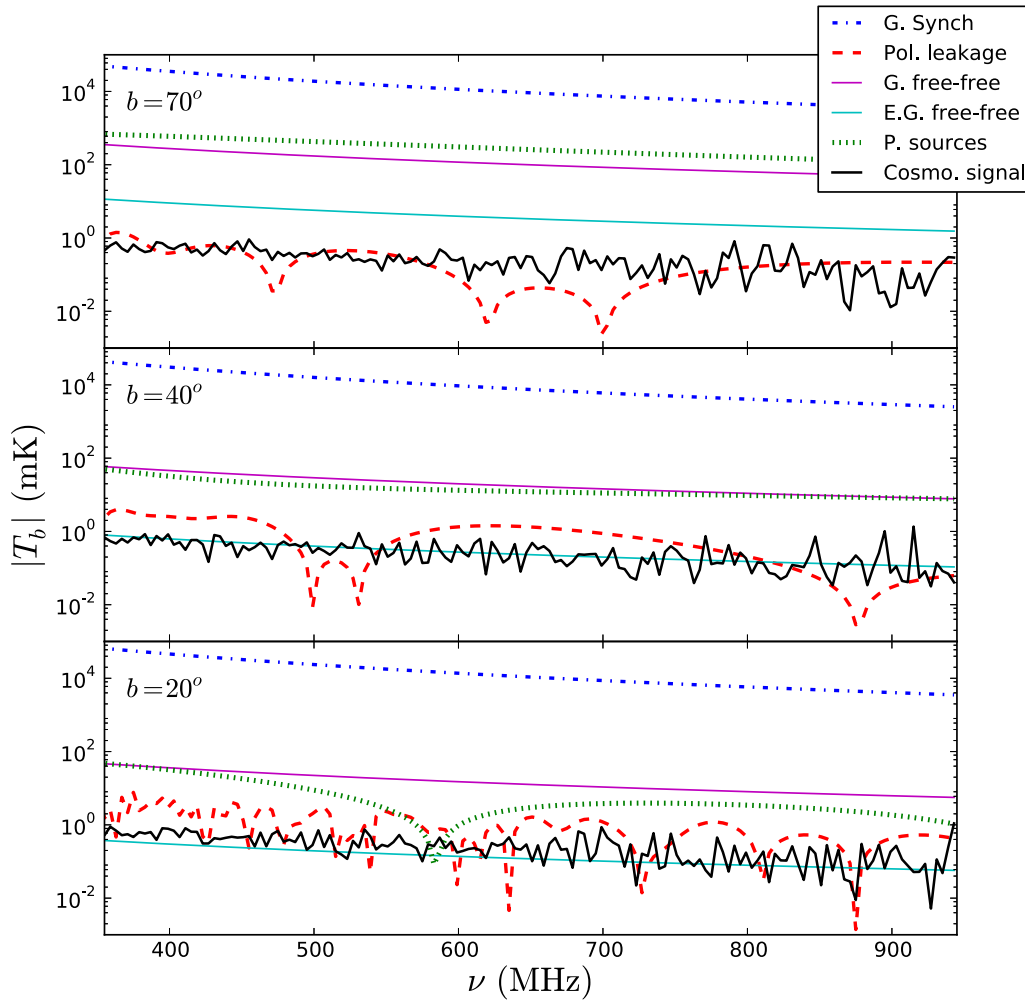
should decrease towards the galactic plane, since the Galaxy becomes ‘thicker’ and the incoherent emission from many regions cancels out. We have computed  $\Pi$  for different frequencies in thick galactic latitude stripes both for the HAMMURABI simulations and for our maps. The result for the fiducial simulation is shown in Fig. 10. As mentioned in Section 4.2.3, this larger depolarization towards the galactic plane is automatically taken into account in our maps through the larger values of  $\sigma_\psi$  computed from the map of Oppermann et al. (2012, see the bottom-right plot in Fig. 6).

## 6 CONCLUSIONS

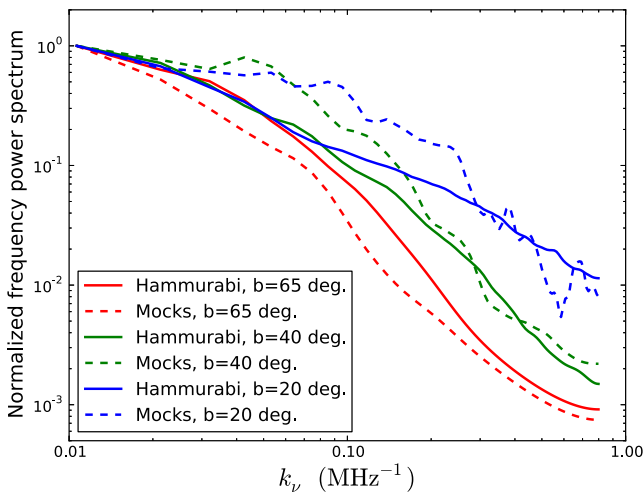
Intensity mapping of H I is potentially a very powerful observational tool to study the large-scale structure of the Universe at late times. In this paper, we have presented a public code that can be used to generate simple and fast mock intensity mapping observations including both the cosmological signal and different galactic and extragalactic foregrounds.

We have verified that our methods are able to reproduce the most important features of these observations. In particular, we have verified that the simulated cosmological signal follows the correct clustering statistics to second order both in the angular directions and along the LOS. Furthermore, we have verified that our models





**Figure 8.** Frequency dependence of the different foregrounds and the cosmological signal along LOS with different galactic latitudes (given in the top-right corner of each panel). The effect of Faraday decorrelation increases as we approach the galactic plane, making the subtraction of the polarization leakage more challenging.



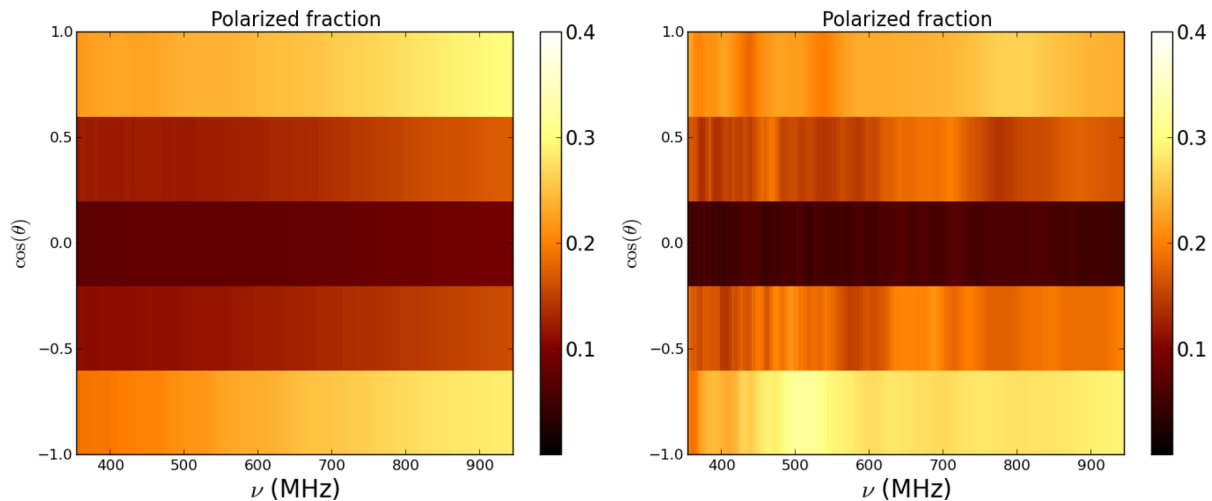
**Figure 9.** Frequency power spectrum at different galactic latitudes for HAMMURABI (solid lines) and for the code presented in this paper (dashes). The curves are normalized to 1 for the first non-zero wavenumber.

for the different foregrounds mimic the most relevant observational effects, including their angular distribution and frequency decorrelation (particularly in the case of the polarized galactic synchrotron, which is probably the most troublesome foreground). Using our code, one full simulation including cosmological signal and foregrounds with a spatial resolution of  $\sim 4 \text{ Mpc } h^{-1}$  and a frequency bandwidth of  $\sim 4 \text{ MHz}$  can be generated in around 30 min using an 80-core machine – ideally we wish to speed this up by a factor of 10 so as to be useful in Monte Carlo analysis.

We believe that these tools could be especially useful for a number of studies, such as

- (i) a thorough evaluation of different foreground-subtraction techniques;
- (ii) optimizing the instrument design to suppress polarization leakage to the appropriate level;
- (iii) assessing the viability of different cosmological analyses, such as BAO, RSDs or primordial non-Gaussianity, in a realistic way (i.e. fully taking into account the effects of foregrounds subtraction);
- (iv) studying the impact of different systematic effects on real observations; and
- (v) a complete statistical analysis of the uncertainties for different experiments.





**Figure 10.** Polarized fraction on different galactic latitudes and frequencies as reproduced by HAMMURABI (left-hand panel) and our own code (right-hand panel).

As has been pointed out in the text, the methods used by our code have some shortcomings that we would like to address in future versions.

(i) We would like to implement alternative methods, such as higher order Lagrangian perturbation theory or quick particle-mesh algorithms, to generate the 3D matter density field. This would be necessary in order to explore higher order moments of the clustering statistics.

(ii) It would be desirable to have a more direct handle on the physical description of galactic synchrotron emission, which would require a  $3D+\nu$  LOS integrator such as HAMMURABI.

(iii) In its current version, our code only supports the simulation of  $w$ CDM models. In the future, we plan to implement other non-standard cosmologies.

(iv) The first public version of the code generates full-sky maps for both the signal and the foregrounds. We are currently working on a future version that will allow the user to simulate smaller patches of the sky and use the flat-sky approximation. This would also imply a major boost in speed and memory usage.

Our code is publicly available and can be found in <http://intensitymapping.physics.ox.ac.uk/CRIME.html>. The code consists of a number of independent subroutines, written in C and FORTRAN 90, that generate the different components. Even though the code has been optimized to use a minimum amount of computer memory, intensity mapping provides naturally an immense amount of data (equivalent to one CMB map per frequency bin), and therefore these simulations can be expensive in terms of memory. For this reason, some users may be more interested in obtaining the simulated maps used for this work rather than generating them. These maps can be found in the same URL.

## ACKNOWLEDGEMENTS

We would like to thank Tessa Baker, Gianni Bernardi, Philip Bull, Tzu-Ching Chang, Clive Dickinson, Matt Jarvis, Thibaut Louis, Sigurd Næss and Jonathan Sievers for useful comments and discussions. The points raised by the journal referee also improved the quality of this work, and we would like to acknowledge his contribution here. DA is supported by ERC grant 259505 and acknowledges the hospitality of the University of the Western Cape. PGF

acknowledges support from Leverhulme, STFC, BIPAC and the Oxford Martin School. MGS acknowledges support from the National Research Foundation (NRF, South Africa), the South African Square Kilometre Array Project and FCT under grant PTDC/FIS-AST/2194/2012.

## REFERENCES

- Abbott T. et al. (Dark Energy Survey Collaboration), 2005, preprint ([astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Abdalla F. B., Rawlings S., 2005, MNRAS, 360, 27
- Anderson L. et al., 2014, MNRAS, 441, 24
- Ansari R. et al., 2012, A&A, 540, A129
- Bagla J., Khandai N., Datta K. K., 2010, MNRAS, 407, 567
- Barkana R., Loeb A., 2005, ApJ, 624, L65
- Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339
- Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239
- Bernardi G. et al., 2009, A&A, 500, 965
- Bernardi G. et al., 2010, A&A, 522, A67
- Beutler F. et al., 2011, MNRAS, 416, 3017
- Blake C. et al., 2011, MNRAS, 415, 2892
- Bull P., Ferreira P. G., Patel P., Santos M. G., 2014, preprint ([arXiv:1405.1452](https://arxiv.org/abs/1405.1452))
- Camera S., Santos M. G., Ferreira P. G., Ferramacho L., 2013, Phys. Rev. Lett., 111, 171302
- Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
- Cole S. et al., 2005, MNRAS, 362, 505
- Coles P., Jones B., 1991, MNRAS, 248, 1
- Colless M. et al., 2001, MNRAS, 328, 1039
- de la Torre S., Peacock J. A., 2013, MNRAS, 435, 743
- De S., Tashiro H., 2014, Phys. Rev. D, 89, 123002
- Delabrouille J. et al., 2013, A&A, 553, A96
- Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576
- Drinkwater M. et al., 2010, MNRAS, 401, 1429
- Fosalba P., Pan J., Szapudi I., 2005, ApJ, 632, 29
- Gleser L., Nusser A., Benson A. J., 2008, MNRAS, 391, 383
- Gong Y., Chen X., Silva M., Cooray A., Santos M. G., 2011, ApJ, 740, L20
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
- Guha Sarkar T., Mitra S., Majumdar S., Choudhury T. R., 2012, MNRAS, 421, 3570
- Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, A&AS, 47, 1

Haverkorn M., Brown J. C., Gaensler B. M., McClure-Griffiths N. M., 2008, *ApJ*, 680, 362

Hinshaw G. et al., 2013, *ApJS*, 208, 19

Jelić V. et al., 2008, *MNRAS*, 389, 1319

Jelić V., Zaroubi S., Labropoulos P., Bernardi G., de Bruyn A. G., Koopmans L. V. E., 2010, *MNRAS*, 409, 1647

Jones D. et al., 2004, *MNRAS*, 355, 747

Kaiser N., 1987, *MNRAS*, 227, 1

Kaiser N. et al., 2002, *Proc. SPIE – Int. Soc. Opt. Eng.*, 4836, 154

Kitaura F.-S., Jasche J., Metcalf R. B., 2010, *MNRAS*, 403, 589

Kogut A. et al., 2007, *ApJ*, 665, 355

Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473

Lidz A., Furlanetto S. R., Oh S. P., Aguirre J., Chang T.-C., Doré O., Pritchard J. R., 2011, *ApJ*, 741, 70

Liu A., Tegmark M., Bowman J., Hewitt J., Zaldarriaga M., 2009, *MNRAS*, 398, 401

Loeb A., Wyithe J. S. B., 2008, *Phys. Rev. Lett.*, 100, 161301

Manera M. et al., 2013, *MNRAS*, 428, 1036

Mao Y., Tegmark M., McQuinn M., Zaldarriaga M., Zahn O., 2008, *Phys. Rev. D*, 78, 023529

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, *ApJ*, 653, 815

Moore D. F., Aguirre J. E., Parsons A. R., Jacobs D. C., Pober J. C., 2013, *ApJ*, 769, 154

Morales M. F., Bowman J. D., Hewitt J. N., 2006, *ApJ*, 648, 767

Oh S. P., Mack K. J., 2003, *MNRAS*, 346, 871

Oppermann N. et al., 2012, *A&A*, 542, A93

Pacholczyk A. G., 1970, *Radio Astrophysics: Nonthermal Processes in Galactic and Extragalactic Sources*. Freeman & Co., San Francisco

Peebles P., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ

Peterson J. B. et al., 2009, in *astro2010: The Astronomy and Astrophysics Decadal Survey*, Vol. 2010, 21-cm Intensity Mapping, p. 234, preprint ([arXiv:0902.3091](https://arxiv.org/abs/0902.3091))

Planck Collaboration, 2013a, preprint ([arXiv:1303.5062](https://arxiv.org/abs/1303.5062))

Planck Collaboration, 2013b, preprint ([arXiv:1303.5076](https://arxiv.org/abs/1303.5076))

Pritchard J. R., Loeb A., 2008, *Phys. Rev. D*, 78, 103511

Rybicki G. B., Lightman A. P., 1986, *Radiative Processes in Astrophysics*. Wiley-VCH, New York

Santos M. G., Cooray A., Knox L., 2005, *ApJ*, 625, 575 (SCK)

Seo H.-J., Dodelson S., Marriner J., McGinnis D., Stebbins A., Stoughton C., Vallinotto A., 2010, *ApJ*, 721, 164

Shaw J. R., Sigurdson K., Pen U.-L., Stebbins A., Sitwell M., 2013, *ApJ*, 781, 57

Shaw J. R., Sigurdson K., Sitwell M., Stebbins A., Pen U.-L., 2014, preprint ([arXiv:1401.2095](https://arxiv.org/abs/1401.2095))

Smith R. E. et al., 2003, *MNRAS*, 341, 1311

Sun X. H., Reich W., Waelkens A., Enßlin T. A., 2008, *A&A*, 477, 573

Switzer E. R. et al., 2013, *MNRAS*, 434, L46

Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, *J. Cosmol. Astropart. Phys.*, 6, 36

Testori J. C., Reich P., Reich W., 2008, *A&A*, 484, 733

Tyson J., 2002, *Proc. SPIE – Int. Soc. Opt. Eng.*, 4836, 10

Waelkens A., Jaffe T., Reinecke M., Kitaura F. S., Enßlin T. A., 2009, *A&A*, 495, 697

Wang X., Tegmark M., Santos M. G., Knox L., 2006, *ApJ*, 650, 529

White M., Tinker J. L., McBride C. K., 2014, *MNRAS*, 437, 2594

Wilman R. J. et al., 2008, *MNRAS*, 388, 1335

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, *ARA&A*, 43, 861

Wolleben M., Landecker T. L., Reich W., Wielebinski R., 2006, *A&A*, 448, 411

Wolleben M. et al., 2009, in *Strassmeier K. G., Kosovichev A. G., Beckman J. E., eds, Proc. IAU Symp. 259, Cosmic Magnetic Fields: From Planets, to Stars and Galaxies*. Cambridge Univ. Press, Cambridge, p. 89

Wolz L., Abdalla F. B., Blake C., Shaw J. R., Chapman E., Rawlings S., 2014, *MNRAS*, 441, 3271

Wyithe J. S. B., Loeb A., 2008, *MNRAS*, 383, 606

Wyithe J. S. B., Loeb A., Geil P. M., 2008, *MNRAS*, 383, 1195

Zel'dovich Y. B., 1970, *A&A*, 5, 84

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, *MNRAS*, 359, L30

## APPENDIX A: VALIDITY OF THE METHOD

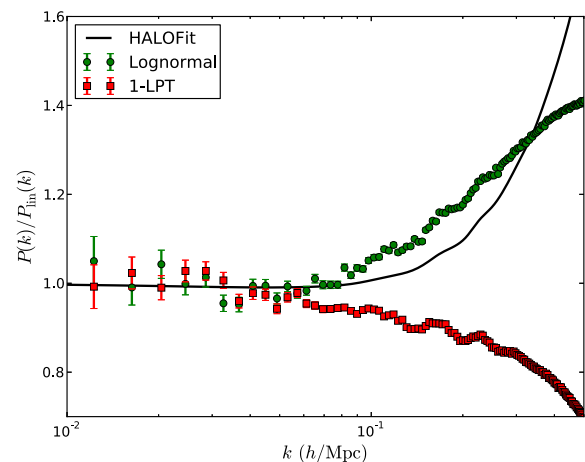
As stated in Section 1, in order to be able to generate fast mock observations it is often necessary to use approximate methods that reduce the computational complexity but that can lead to quantitatively incorrect results. This is the case of the method presented here, and therefore we would like to address the validity of some of these approximations here.

### A1 Non-linearities

Even though small non-linear scales are not particularly relevant for intensity mapping, due to the low angular resolution of most experiments (Bull et al. 2014), it is still interesting to analyse the regime in which the lognormal approximation described in Section 2.2 yields valid results. We have done so by generating a realization of the density field in a comoving box at a fixed time (i.e. no lightcone evolution is applied). We have computed the power spectrum for this realization and compared it with the theoretical linear power spectrum provided by CAMB and its non-linear prediction modelled by HALOFIT (Smith et al. 2003).

We have generated a lognormal realization of the density field at redshift  $z = 0.7$  using the same cosmological parameters and spatial resolution that was used for the ACCURATE simulation (see Table 2). Furthermore, in order to compare the validity of this method with the results for other techniques used in the literature (Manera et al. 2013), we have also generated the density field as predicted by first-order Lagrangian perturbation theory [1LPT – i.e. the Zel'dovich approximation (Zel'dovich 1970)]. For this, we have generated the first-order Lagrangian displacements for a set of comoving tracers of the density field at the same redshifts (this is equivalent to setting the initial conditions for an  $N$ -body simulation at  $z_{\text{init}} = 0.7$ ). We then computed the density field by cloud-in-cell interpolation of the positions of these tracers.

Fig. A1 shows the ratio of the power spectra of these two realizations to the linear power spectrum (green circles for the lognormal



**Figure A1.** Ratio of the power spectrum computed from simulations to the linear power spectrum. The simulated density fields were generated in at a fixed time corresponding to  $z = 0.7$  in a box with the same spatial resolution as was used for the ACCURATE simulation. We show results for a lognormal density field (green circles), a 1LPT field (red squares) and the non-linear prediction given by HALOFIT (black solid line).

field and red squares for 1LPT), as well as the same ratio for the non-linear power spectrum (black line). It is evident from this figure that the lognormal transformation is not able to perfectly reproduce the non-linear power spectrum on all scales, although it does provide extra power with respect to the linear power spectrum for large  $k$ . The same is true for 1LPT: shell-crossing at low redshift smoothes the density field, and all power is lost on small scales. We can thus conclude that it is safe to use the method described in this work within the linear regime, which is the most relevant one for intensity mapping. However, any analysis requiring a good description of the non-linear clustering should make use of a more sophisticated method.

## APPENDIX B: SCALE-DEPENDENT BIAS

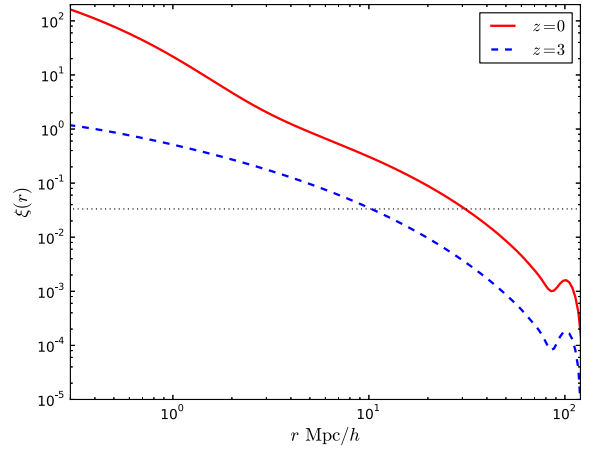
Our code is able to include an evolving linear bias  $b_{\text{H1}}(z)$  that is constant for all scales. As is the case for most tracers of the density field, this can only be a good approximation for large scales, and it is important to know the regime in which scale dependence becomes relevant. This has been studied in the literature, and in particular Guha Sarkar et al. (2012) give constraints on both scale dependence and redshift evolution based on results from  $N$ -body simulations. Their results show that  $b_{\text{H1}}$  is constant in scale to a good approximation on scales  $k \lesssim 0.5 h \text{ Mpc}^{-1}$  for  $z < 3$ .

### B1 Higher order correlations

Since lognormal density fields are based on a local transformation of a Gaussian field, it is a well-documented fact that they are unable to reproduce the three-, four- and higher-order correlations produced by the non-linear gravitational collapse (White et al. 2014). This is not a problem for any science case that relies only on the two-point function, as is usually the case; however, we can estimate the scale at which higher order correlations become relevant.

From the study of the three-point correlation function  $\zeta$  in  $N$ -body simulations, we know that it is possible to parametrize it as

$$\zeta(r_1, r_2, r_3) = Q_3 [\xi(r_1)\xi(r_2) + \xi(r_2)\xi(r_3) + \xi(r_3)\xi(r_1)],$$



**Figure A2.** Non-linear two-point correlation function at  $z = 0$  (solid red) and  $z = 3$  (dashed blue). The intercepts of these curves with the horizontal dotted line show the approximate scale at which higher order correlations are comparable to  $\xi(r)$ .

where  $Q_3$  is a function of the three relative distances  $r_i$  that takes values  $Q_3 \sim \mathcal{O}(1-3)$  (Peebles 1980; Fosalba, Pan & Szapudi 2005). We can thus get an idea of the scale at which the amplitude of  $\zeta$  is comparable to that of the two-point correlation  $\xi$  by focusing on the equilateral configuration,  $r_i = r$ , in which case

$$\frac{\zeta(r, r, r)}{\xi(r)} \sim \mathcal{O}(1-3) 3 \xi(r). \quad (\text{B1})$$

By choosing an arbitrary ratio of  $\zeta/\xi = 0.1$ , we can then say that higher order correlations become important when  $3\xi(r) \sim 0.1$ . Between redshifts  $z = 0$  and 3, this transition occurs in the range  $r \sim 30-10 \text{ Mpc } h^{-1}$  (see Fig. A2).

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.